

Bayesian Inference for Color Image Quantization via Model-Based Clustering Trees

Fionn Murtagh, Adrian E. Raftery, Jean-Luc Starck ¹

Technical Report no. 402
Department of Statistics
University of Washington

November 2, 2001

¹F. Murtagh is with the School of Computer Science, Queen's University Belfast Belfast BT7 1NN, Northern Ireland (e-mail f.murtagh@qub.ac.uk). A.E. Raftery is with the Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322 (e-mail raftery@stat.washington.edu). J.-L. Starck is with SEI-SAP/DAPNIA, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France (e-mail jstarck@cea.fr).

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 02 NOV 2001		2. REPORT TYPE		3. DATES COVERED 00-11-2001 to 00-11-2001	
4. TITLE AND SUBTITLE Bayesian Inference for Color Image Quantization via Model-Based Clustering Trees			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Washington, Department of Statistics, Box 354322, Seattle, WA, 98195-4322			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 22	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Abstract

We consider the problem of color image quantization, or clustering of the color space. We propose a new methodology for doing this, called model-based clustering trees. This is grounded in model-based clustering, which bases inference on finite mixture models estimated by maximum likelihood using the EM algorithm, and automatically chooses the number of clusters by Bayesian model selection, approximated using BIC, the Bayesian Information Criterion. We build a clustering tree by first clustering the first color band, then using the second color band to cluster each of the clusters found at the first stage, and the resulting clusters are then further subdivided in the same way using the third color band. The tree is pruned automatically as part of the algorithm by using Bayesian model selection to choose the number of clusters at each stage. An efficient algorithm for implementing the methodology is proposed. The method is applied to several real data sets and compared, with good results, to an alternative method that clusters simultaneously on all bands.

KEYWORDS: Bayesian modeling; Clustering; Color-space quantization.

Contents

1	Introduction	1
2	Model-Based Clustering Trees	2
2.1	Univariate Normal Finite Mixture Models	2
2.2	Choosing the Number of Clusters via Bayesian Model Selection	4
2.3	Model-Based Clustering Trees Algorithm	5
3	Algorithm Design	6
3.1	Algorithm Properties	6
3.2	Unconstrained Band Ordering	7
3.3	Transformations	8
4	Examples	9
4.1	Fitting Model-Based Clustering Trees	9
4.2	Comparison with ImageMagick: Separable Band Quantization	10
5	Discussion	11

List of Tables

1	Cardinalities of embedded partitions for successive bands, 1, 2 and 3. Sail data.	9
---	--	---

List of Figures

1	Tulips test image, original data. The original is in color, and this is shown here in greyscale.	12
2	Quantization using our algorithm. Number of quantization levels in R, G and B: 9, 9, 8. The original is in color, and this is shown here in greyscale.	13
3	Quantization using ImageMagick. Number of quantization levels in R, G and B: 8, 9, 8. The original is in color, and this is shown here in greyscale.	14
4	Sails test image, original data. The original is in color, and this is shown here in greyscale.	15
5	Quantization using our algorithm. Number of quantization levels in R, G and B: 9, 8, 9. The original is in color, and this is shown here in greyscale.	16
6	Quantization using ImageMagick. Number of quantization levels used in R, G and B: 9. The original is in color, and this is shown here in greyscale. Dithering is used, yielding a visually better result than no dithering.	17

1 Introduction

Clustering and segmentation in an image analysis context have a long history [23]. Objectives include: quantization of data values for later use with a codebook in a compression context; targeting delivery to display devices supporting small, bounded pixel data value depth; as a preliminary to object and feature detection and analysis in images; and as a basis for other image processing operations such as image registration and archiving.

Color quantization (see, e.g., [33]) is unsupervised classification, or clustering, of the color space. Typically a potential 256^3 (for 1-byte pixel values in each color band) distinct colors are to be mapped into a reduced number of palette colors, e.g. up to 256 [27]. Color image quantization could be said to be brittle because the distribution of local optima in the 3-dimensional color space is expected to be broad [28].

In this work we will be concerned with 3-band color data. However our results will carry over to multiple band data (e.g. in Earth observation), or to multi- or hyperspectral data in any domain. Our work may similarly be relevant for aspects of image sequence and video analysis.

A short review of software for quantization is as follows. In the image display and analysis program xv, Bradley [3] sequentially picks colors which are as unlike as possible to all already picked colors. RGB space is used. In ImageMagick, Cristy [8] determines a color description tree, associates the frequency of occurrence with each node in this tree, and a quantization error defined in RGB space. ImageMagick then prunes the tree based on these measures. An algorithm developed by Wan et al. [35] also uses hyperboxes with hyperplane splits, based on a sum of squares error criterion.

In this paper, we propose a new method for color image quantization, called *model-based clustering trees*. This combines maximum likelihood estimation of finite mixture models with Bayesian model selection. It operates recursively on the three color bands of the image. First it clusters the pixels on the basis of the first band. Then, using the second color band, it clusters each of the clusters found in the first stage. The third stage further subdivides the resulting clusters on the basis of the third band. Bayesian model selection is used at each stage to determine the number of clusters, so that the data are used to decide adaptively the extent to which the tree is pruned.

The resulting method allows the number of color quantization levels to be chosen on the basis of the data, if desired. If the number of quantization levels is predetermined, however [24] the method can easily handle this as a special case. Also, a natural order often exists on

color band dimensions, e.g., chromaticities convey far less perceptual information than does the luminosity (see, e.g., [36]). Such an order can be readily accommodated in our approach. Experimental results are presented below for the cases of an order on color bands, and the case of no such order.

We can readily accommodate noise in our image data. This is implied by image features taken as realizations of distributional models. Explicit noise components are readily incorporated into our modeling, as discussed in earlier work of ours [4]. Our MR software package [20] provides color image noise filtering, together with compression, functionality.

We can accommodate a very small number of classes (clusters or segments) for the pixels, or a large number. A small number of classes may be needed as a preliminary to a data interpretation, or high-level vision stage of the analysis, or to cater for 4-bit displays, in which case optimal 16-level quantization is targeted. A large number of classes may be needed when high fidelity to the original image is required, by close approximation to 8-bit or greater precision of pixel data in each color band.

In Section 2 we describe the model-based clustering trees methodology. In Section 3 we discuss aspects of algorithm design and properties. In Section 4 we show how the algorithm performs on several real datasets, and compare it to an alternative algorithm that clusters separately (rather than recursively) on the color bands.

2 Model-Based Clustering Trees

Our basic framework is that of *model-based clustering*, as described, for example, by Fraley and Raftery [14, 15]. In this methodology, a finite mixture of normal distributions is fit to the data by maximum likelihood estimation using the EM algorithm, the number of groups is chosen using Bayesian model selection, and if hard clustering is desired, each datum is assigned to its most likely group *a posteriori*. Model-based clustering *trees* produces a clustering of multivariate data by clustering on each band or dimension recursively.

We now briefly outline finite mixture modeling, Bayesian model selection, and model-based clustering trees.

2.1 Univariate Normal Finite Mixture Models

In the univariate normal finite mixture model, one-dimensional observations x_i are assumed to be drawn from G groups, each of which is normally distributed. The g -th group has mean μ_g and variance σ_g^2 . Given observations $x = (x_1, \dots, x_n)$, let γ be an unobserved $n \times G$ cluster

assignment matrix, where $\gamma_{ig} = 1$ if x_i comes from the g -th group, and $\gamma_{ig} = 0$ otherwise. Our goals are to determine the number of clusters G , to determine the cluster assignment of each datum, and to estimate the parameters μ_g and σ_g of each cluster.

The probability density for this model is

$$f(x_i|\theta, \lambda) = \sum_{g=1}^G \lambda_g f_g(x_i|\theta_g), \quad (1)$$

where $\theta_g = (\mu_g, \sigma_g^2)^T$, $f_g(\cdot|\theta_g)$ is a normal density with mean μ_g and variance σ_g^2 , $\theta = (\theta_1, \dots, \theta_G)$, and $\lambda = (\lambda_1, \dots, \lambda_G)$ is a vector of mixture probabilities such that $\lambda_g \geq 0$ ($g = 1, \dots, G$) and $\sum_{g=1}^G \lambda_g = 1$.

We estimate the parameters by maximum likelihood using the EM (expectation-maximization) algorithm [10, 19], one of the most successful methods in modern statistics. For its application to model-based clustering, see [18, 6, 9]. This is a procedure for iteratively maximizing likelihoods in situations where there are unobserved quantities and estimation would be simple if these were known. In the clustering case, the unobserved quantities are the cluster assignments given by the matrix γ .

The EM algorithm iterates between the E step and the M step. In the E step, the conditional expectation, $\hat{\gamma}$, of γ given the data and the current estimates of θ and λ is computed, so that $\hat{\gamma}_{ig}$ is the conditional probability that x_i belongs to the g -th group. In the M step, conditional maximum likelihood estimators of θ and λ given the current $\hat{\gamma}$ are computed.

The point is that the E step and the M step are both simple, so that the EM algorithm as a whole is also simple. By contrast, direct maximization of the likelihood for the mixture model is complex in general. Although the EM algorithm has some limitations (e.g. it is not guaranteed to converge to a global rather than a local maximum of the likelihood), it is generally efficient and effective for Gaussian clustering problems. The EM algorithm requires a starting point. An initial classification for each possible number of clusters may be found by agglomerative hierarchical model-based clustering [22, 2].

This procedure is especially efficient for clustering image pixels using single color bands or greyscale images. In general the EM algorithm requires $O(n)$ time, where n is the number of pixels. However, typically pixels can have one of only a limited number, ℓ , of intensities in each band, such as 256. If we first summarize the data by the counts of the numbers of pixels with each intensity level, the EM algorithm becomes an $O(\ell)$ algorithm rather than an $O(n)$ one. Since n is often on the order of 65,000 or more, and ℓ is typically 256, this is a major speed-up and provides a compelling reason for clustering one band at a time if

this can be done without degrading performance too much. This is the motivation for the model-based clustering trees methodology described in Section 2.3.

2.2 Choosing the Number of Clusters via Bayesian Model Selection

We use Bayesian model selection to choose the number of clusters. For review of Bayesian model selection, see Kass and Raftery [16] and Raftery [25]. Pioneering work in this area was due to H. Jeffreys, I.J. Good and (according to the latter) A. Turing.

We consider a range of candidate numbers of clusters, $G = G_{\min}, \dots, G_{\max}$. Each possible number of clusters, G , implies a different statistical model for the data, M_G . The model M_G has a vector of unknown parameters, ψ_G , consisting of the G means, the G variances, and the $(G-1)$ independently estimated mixture probabilities: $(3G-1)$ parameters in all. Our prior model probabilities are $p(M_G)$ for $G = G_{\min}, \dots, G_{\max}$, where $\sum_{G=G_{\min}}^{G_{\max}} p(M_G) = 1$. Often each number of clusters considered is taken to be equally likely *a priori*, so that $p(M_G) = 1/(G_{\max} - G_{\min} + 1)$ for each G . The model parameters ψ_G also have prior distributions $p(\psi_G|M_G)$, which are typically rather diffuse and do not affect the final conclusions unduly. The data produce posterior model probabilities, $p(M_G|x)$, where again $\sum_{G=G_{\min}}^{G_{\max}} p(M_G|x) = 1$.

By Bayes' theorem,

$$p(M_G|x) = \frac{p(x|M_G)p(M_G)}{\sum_{H=G_{\min}}^{G_{\max}} p(x|M_H)p(M_H)}, \quad G = G_{\min}, \dots, G_{\max}. \quad (2)$$

In (2), $p(x|M_G)$ is the *integrated likelihood* of model M_G , which requires integration over the model's parameter space, as follows:

$$p(x|M_G) = \int p(x|\psi_G, M_G)p(\psi_G|M_G)d\psi_G, \quad (3)$$

by the law of total probability.

The integral (3) is intractable analytically and is not easy to evaluate. However, twice the logarithm of the integrated likelihood can be approximated by the Bayesian Information Criterion, or BIC:

$$\begin{aligned} 2 \log p(x|M_G) &\sim 2 \log p(x|\hat{\psi}_G, M_G) - (3G-1) \log n \\ &= \text{BIC} \end{aligned} \quad (4)$$

(Schwarz 1978; Kass and Wasserman 1995; Raftery 1995). In (4),

$$p(x|\hat{\psi}_G, M_G) = \prod_{i=1}^n \sum_{g=1}^G \hat{\lambda}_g f_g(x_i|\hat{\theta}_g)$$

is the maximized likelihood. In words, $BIC = 2(\log \text{maximized likelihood}) + (\log n)(\text{number of parameters})$. The BIC measures the balance between the improvement in the likelihood and the number of model parameters needed to achieve that likelihood. While the absolute value of the BIC is not informative, differences between the BIC values for two competing models provide estimates of the evidence in the data for one model against another. The use of the BIC in choosing clusters in a mixture or clustering model is discussed by Roeder and Wasserman [26] and Dasgupta and Raftery [9]. Applications are in Campbell et al. [4, 5], Mukherjee et al. [21] and in other articles.

2.3 Model-Based Clustering Trees Algorithm

For color image segmentation, we cluster the first band, then we cluster each of the resulting clusters using the second band, and finally we subdivide the resulting clusters yet again using the third band. This requires choosing the 3D color space and the order in which the bands are to be clustered; these issues are discussed in Section 3.

The algorithm can be summarized as follows:

1. For the first color band, use BIC to choose the number of clusters (here we use $G_{\min} = 1$, $G_{\max} = 9$). Use the EM algorithm to estimate the parameters of the mixture model, and assign each pixel to the group to which it is most likely to belong *a posteriori*.
2. For each cluster identified in step 1, carry out a separate model-based cluster analysis, this time using only the pixel intensities in the second color band. Each cluster identified in step 1 is then itself subdivided into several clusters.
3. For each (sub)cluster identified in step 2, subdivide it further using the same procedure as in step 2, but this time using only the pixel intensities in the third color band.

The univariate normal mixture model fitting was carried out using an algorithm initially developed by Stanford [30]. In implementing the algorithm, numeric sub-segment mean values are down-scaled by a factor of 10 (i.e., cluster labels expressed as a sequence number 1, 2, ... , and subsequently divided by 10) for level 2 sub-segments, and by a factor of 100 (i.e., cluster labels expressed as a sequence number 1, 2, ... , and subsequently divided by 100) for level 3 sub-segments. Given the linear order properties explored in the next section, this facilitates addition of segment and sub-segment means. It is clear that this algorithm provides an adaptive stepwise procedure for pruning a segment tree with an upper limit of $9^3 = 729$ segments.

Timing tests were carried out in order to look at scaling with respect to number of clusters, and image dimensions. Scaling with respect to number of clusters is approximately linear. For 3, 9, 15, 21, 27, 33 and 39 clusters from a 768×512 array, the compute times were, respectively, 20, 40, 53, 69, 104, 108, 131 seconds. Scaling with input image dimensionality was as follows, where floating point storage was used for input images, and 10 clusters were found in each case: (300×300) , 11 seconds; (600×600) , 35 seconds; (900×900) , 87 seconds. This is better than linear in this dimensionality region.

3 Algorithm Design

3.1 Algorithm Properties

In this section, we review results for:

1. Band segmentation, and spectral clustering.
2. Label monotonicity.
3. Partial order of clusters, and total order of cluster labels.

Contiguity of pixels associated with a cluster implies that segmentation rather than clustering holds. In the spatial or image dimension space domain, our algorithm provides a clustering solution. In color or spectral band space, our algorithm provides a segmentation solution. It is clear that the latter allows for straightforward inducing of a total order relation on the image pixels, which is necessary for display as a single image plane.

In the following we will discuss further properties of our algorithm related to segmentation in the spectral domain.

To facilitate image display, we posit the following label monotonicity principle.

Label monotonicity principle:

For clusters c_i of means $m(c_i)$, and labels $l(c_i)$ we require: $l(c_i) < l(c_j) \iff m(c_i) \prec m(c_j)$ for all i and j .

The relation \prec is a total order on the multivariate means.

Such a monotonicity principle is desirable since the relation \prec can be mapped onto a color lookup table, allowing the labels to be mapped monotonically.

The mixture modeling algorithm is a segmentation algorithm since it is one-dimensional and we have the following exact relationship for any given spatial dimension or color band: $l(c_i) < l(c_j) \iff m(c_i) < m(c_j)$.

Next, consider clusters defined from band or color 2, c_{2i} , and we will write c_{1i} for a cluster defined from band 1. For given c_{1i} , we have $c_{1i} = \cup_{j \in J_i} c_{2j}$ where the cardinality of band 2 cluster set, J_i , is at least 1. When $|J_i| = 1$ we ignore any contribution by band 2, i.e., no sub-clusters have been found.

A fortiori it holds that $l(c_{2i}) < l(c_{2j}) \iff m(c_{2i}) < m(c_{2j})$ for all band 2 segments.

Similar properties hold for band 3. We have $c_{1i} = \cup_{j \in J_i} c_{2j}$ for a set of subclusters, J_i , and $c_{2i'} = \cup_{j' \in J'_i} c_{2j'}$ for a set of subclusters, J'_i , which also is at least of cardinality 1 and at most of cardinality $|c_{2i'}|$. Again we have a total order on all band 3 segments.

Our algorithm produces an embedded set of segments. Is there a total order on segment means resulting from this embedded set, so that the principle of label monotonicity is respected? To show that this is indeed the case, we firstly note that band 3 segments replace a parent band 2 segment, and band 2 segments replace a parent band 1 segment.

Secondly, at band 2, given $c_{1i} = \cup_{j \in J_i} c_{2j}$, we can ensure that $\max_j c_{2j} < c_{1(i+1)}$ since by the total order on band 2 and on band 1 segments, we have a well-defined maximum value of c_{2j} , and similarly $c_{1(i+1)} > c_{1i}$ is next in band 1 sequence.

These two properties ensure that there results a total order of segment means. Hence it follows that the label monotonicity principle is respected.

3.2 Unconstrained Band Ordering

In this section, we follow closely Tate [31] who considers the band ordering problem for compression of multispectral images.

We consider the problem of clustering on one band coordinate, assuming that this band presents good clustering properties, followed by clustering on a second band based on the first band clustering, and so on. As expressed in the previous section, if c_{1i} is the i th cluster from the first band, then we seek clusters c_{2j} such that $c_{1i} = \cup_{j \in J_i} c_{2j}$, i.e., J_i is a partition of cluster c_{1i} . Similarly we proceed to a third band, based on available results for bands 1 and 2.

The order in which we consider the bands is evidently important. Let us define goodness of clustering as the tightness of the clusters, i.e. the minimum sum of variances of clusters for all bands.

Clearly a result of this definition is that when we find no subclustering at bands 2 and

3, the corresponding subcluster variances are equal to 0, and hence the contribution to the overall sum of variances is thereby minimized.

Finding the optimal band ordering for clustering is not difficult in the color space case, since it involves only 6 possible alternatives. In the case of hyperspectral data, with many dozens or hundreds of bands, the situation is more computationally demanding.

Define graph $G = (V, E)$ such that an edge E_{ij} has weight w_{ij} representing the added clustering quality attainable by clustering band i before band j . By design, band i is partitioned, and the clusters of the i -partition are each partitioned based on band j information. Define w_{ij} as the *improvement* in the overall sum of posteriors (or sum of intra-cluster variances: the criterion used to quantify clustering quality is not important here) by taking band i before band j .

The problem of finding an optimal band order in the general case of many bands is equivalent to finding an optimal traveling salesman path in the graph, G . This Hamiltonian path problem is NP-hard and the corresponding decision problem of knowing whether we have or do not have a Hamiltonian path in G is NP-complete.

3.3 Transformations

In this section, we investigate image transformation band ordering

RGB coordinates are converted to luminance-hue-saturation YIQ using the following transformation.

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.28 & -0.32 \\ 0.21 & -0.52 & 0.31 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (5)$$

Cheng et al. [7] present a useful overview of color space transformations (unfortunately marred by many typographical errors including in the definition of the above transformation.)

Luminance alone provides a best monochrome approximation. The I and Q chrominance components can be significantly limited (e.g., using decimation) without noticeable image degradation. This transformation therefore provides for an order on the image bands. So also does an eigen-analysis of the multiband image. Finally, we investigate also a color enhancement procedure proposed in Toet [32]: perform sigma-clipping on each color band, followed by color saturation:

Color space, order	Number of quantization levels
RGB	9, 73, 125
GBR	8, 69, 240
RBG	9, 83, 214
YIQ	9, 77, 218
PC1, PC2, PC3 of RGB	9, 78, 180
RGB clipped, saturated	9, 73, 208

Table 1: Cardinalities of embedded partitions for successive bands, 1, 2 and 3. **Sail** data.

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \frac{C_1}{C_1 - C_2} \begin{pmatrix} R - C_2 \\ G - C_2 \\ B - C_2 \end{pmatrix} \quad (6)$$

where $C_1 = \max\{ R, G, B \}$, and $C_2 = \min\{ R, G, B \}$.

4 Examples

Test data sets from [17] were used. These included the **sail** of dimensions 3, 768, 512; **tulips** of dimensions 3, 767, 512; and **frymire** images.

4.1 Fitting Model-Based Clustering Trees

We begin by partitioning a first band, and then continuing with the partitioning of previous band clusters. By constraining the mixture model to 9 image marginal Gaussian components, this yields a tree of Gaussian components of, at most, $9^3 = 729$ nodes. The BIC model assessment criterion will tell us if we can justifiably retain a pruned version of this tree of Gaussians.

Using the **sails** data, the numbers of quantization levels found for bands 1, 2 and 3 are shown in Table 1. Of note is that there is no experimental advantage in the use of band ordering such as through use of YIQ color space, or eigen-images (principal component images, PC1, PC2, PC3). In all cases, increased numbers of clusters entails corresponding improvements in visual quality, and generally in global measures such as mean square error (MSE) relative to input data. MSE values are given in the next subsection.

4.2 Comparison with ImageMagick: Separable Band Quantization

We now turn to look, for comparison, at an alternative method based on quantization carried out independently on the color bands, leading to a small number of quantization levels on each. The model-based clustering trees algorithm takes a color band, and determines the best quantization. Then in a stepwise procedure, further bands are taken and already existing quantized bands are further quantized. Our algorithm is therefore a progressive quantization one. In ImageMagick, by contrast, the quantization algorithm of Cristy [8] carries out quantization directly in the color space. Such full multivariate quantization ought to be superior but, as we will see, gives results which are less acceptable.

We consider the RGB `sail` example. Our algorithm produced 9 quantized levels on the first band, 73 on the second, and 125 on the third (cf. Table 1). This result is a progressive one, with an approximation to the original image based on 9 quantized levels, a better approximation to the original data based on 73 quantized levels, and finally an approximation to the original data based on 125 quantized levels.

ImageMagick’s quantize procedure was given 125 quantization levels as the target, and produced 91, 89 and 94 quantization levels on bands 1, 2 and 3, respectively (i.e., R, G, and B bands).

Mean square error (MSE) discrepancy between the original data and the quantized data closely reflects the number of quantization levels used in the approximation. For band 1, MSEs for our algorithm (9 quantization levels) and Cristy’s (91 quantization levels) were 124.95 and 99.54. For band 2, MSEs for our algorithm (73 quantization levels) and Cristy’s (89 quantization levels) were 127.66 and 127.54. For band 3, MSEs for our algorithm (125 quantization levels) and Cristy’s (94 quantization levels) were 126.28 and 135.29. Histogram equalization was used prior to these calculations, which was necessary for rescaling given that our coding involved 0.1 and 0.01 color step-size increments for bands 2 and 3.

Considering the RGB `tulips` example, we find for our method 9, 72, and 237 progressively quantized bands. For Cristy [8], we find 113, 133, and 133 quantized levels for the three bands, respectively. The user parameter yielding this result was 237 quantization levels.

Qualitatively, the results are very different. Figures 1, 2 and 3 show the `tulips` image.

Figure 1 is a greyscale version of the 768×512 image.

Figure 2 results from applying our algorithm to the RGB data volume. For comparability with the ImageMagick Cristy [8] result, we took a 9 quantization level result for the R band, a 9 quantization level result for the G band, and an 8 quantization level result for the B

band. (The R band quantization levels were tested using BIC against all smaller numbers of levels, and the G and B bands were tested for 8 and 9 levels.)

Figure 3 results from the Cristy ImageMagick quantization procedure. In the latter, 9 quantization levels were requested, yielding an image with 8, 9, and 8 quantization levels for the three color bands. Taking the R band alone of the ImageMagick result, the good fit to the input data (MSE 90.13 for Cristy as opposed to 128.54 for our algorithm, R band in both cases) was explained by a somewhat crude quantization rendition of the image. A better rendition of the image is obtained by straightforwardly converting the quantized 3D data volume, with 8, 9, and 8 quantization levels, to greyscale. This is shown in Figure 3.

Even allowing for the degradation inherent in down-sizing and converting to greyscale Postscript, we find, with reference to the original data in Figure 1, that the output of our algorithm in Figure 2 is superior to the ImageMagick result in Figure 3. The ImageMagick result is however faster than our current implementation. On a Sun Sparc 10 workstation, it takes approximately 10 seconds for the `tulips` data, compared to 40 seconds for each of the three color bands in the case of our algorithm.

Figures 4, 5 and 6 show the `sails` data. Considerable visual detail retained by our algorithm (Figure 5) is lost in the ImageMagick quantization (Figure 6).

5 Discussion

We have shown how a Bayesian modeling approach, model-based clustering trees, can lead to good results in the area of color image clustering. Formal underpinnings for such an algorithm facilitate choice of system parameters (e.g. number of clusters) which in a general setting would be set arbitrarily.

The fitting of a tree of Gaussian components may be of benefit in the exploration of general parameter spaces, since we are not overly dependent on an analysis function or kernel (here: Gaussian) of given morphology. The partial order resulting from the tree of Gaussian components can (at least in principle) accommodate arbitrary alignments or curves in multidimensional clusters. Djorgovski et al. [11] describe a burgeoning need for solving such problems.

The two algorithms studied in this work – based on fitting of a tree of Gaussians, and separable clustering – are of potential interest also for other multiple band, and indeed general multidimensional, data sets. Future work will be on these themes.

One possible improvement on the work would be to prune the tree further using hierar-



Figure 1: Tulips test image, original data. The original is in color, and this is shown here in greyscale.

chical model-based clustering [2]. Here one would start with the partition resulting from our method, treat that as an initial partition, and perform agglomerative hierarchical clustering. One would use the loglikelihood as the clustering criterion for choosing the groups to be merged at each stage, and BIC as the criterion for deciding when to stop merging. This could yield a result with fewer quantization levels, but without degrading the quality of the quantization.

An alternative approach is to carry out model-based clustering using mixtures of multivariate normal distributions directly on the three-dimensional pixel intensities. This was done for MRI images [12], using images with more than three bands. It is highly time-consuming however, and in the MRI work it was feasible only because many pixels could be excluded *a priori*. Our approach is much more efficient computationally and, from the examples described, captures most forms of multivariate clustering in color space very well. It is possible to construct situations where our method would perform much less well than full



Figure 2: Quantization using our algorithm. Number of quantization levels in R, G and B: 9, 9, 8. The original is in color, and this is shown here in greyscale.

multivariate model-based clustering, but we have not encountered such situations in practice, and they would seem rather contrived. Nevertheless, if it were possible to implement full multivariate model-based clustering on large numbers of pixels in comparable compute time to our method, this would seem worth doing, but this has not yet been accomplished.

Acknowledgments

A.E. Raftery's research was supported by ONR grant N00014-96-1-0192. He thanks Chris Fraley and Derek Stanford for helpful discussions.

References

- [1] S. Banerjee and A. Rosenfeld, "Model-based cluster analysis", *Pattern Recognition*, 26, 963–974, 1993.

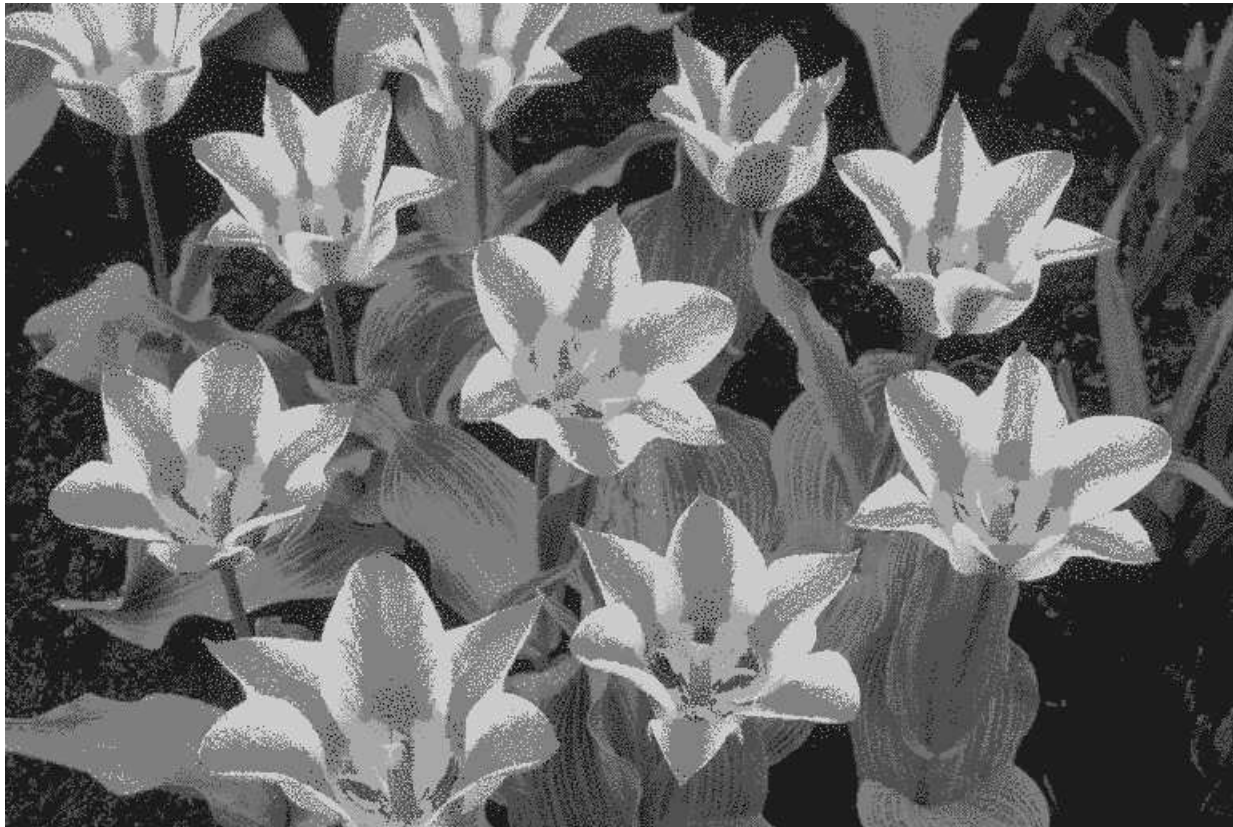


Figure 3: Quantization using ImageMagick. Number of quantization levels in R, G and B: 8, 9, 8. The original is in color, and this is shown here in greyscale.

- [2] J.D. Banfield and A.E. Raftery, “Model-based Gaussian and non-Gaussian clustering”, *Biometrics*, 49, 803–821, 1993.
- [3] J. Bradley, XV – Interactive Image Display for the X Window System, Appendix F, “The diversity algorithm”, 1994. Available at <http://www.trilon.com/xv/>
- [4] J.G. Campbell, C. Fraley, F. Murtagh and A.E. Raftery, “Linear flaw detection in woven textiles using model-based clustering”, *Pattern Recognition Letters*, 18, 1539-1548, 1997.
- [5] J.G. Campbell, C. Fraley, D. Stanford, F. Murtagh and A.E. Raftery, “Model-based methods for textile fault detection”, *International Journal of Imaging Science and Technology*, 10, 339-346, 1999.
- [6] G. Celeux and G. Govaert, “Gaussian parsimonious clustering models”, *Pattern Recognition*, 28, 781–793, 1995.



Figure 4: **Sails** test image, original data. The original is in color, and this is shown here in greyscale.

- [7] H.D. Cheng, X.H. Jiang, Y. Sun and Jingli Wang, “Color image segmentation: advances and prospects”, *Pattern Recognition*, 34, 2259–2281, 2001.
- [8] J. Cristy, ImageMagick routine “Quantize”, <http://www.imagemagick.org/www/quantize.html>, 2001
- [9] A. Dasgupta and A.E. Raftery, “Detecting features in spatial point processes with clutter via model-based clustering”, *Journal of the American Statistical Association*, 93, 294–302, 1998.
- [10] A.P. Dempster, N.M. Laird and D.B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society, Series, B*, 39, 1–22, 1977.
- [11] S.G. Djorgovski, A. Mahabal, R. Brunner, R. Williams, R. Granat, D. Curkendall, J. Jacob and P. Stolorz, “Exploration of parameter spaces in a virtual observatory”, in J.L.



Figure 5: Quantization using our algorithm. Number of quantization levels in R, G and B: 9, 8, 9. The original is in color, and this is shown here in greyscale.

Starck and F. Murtagh, Eds., *SPIE Proceedings* Vol. 4472, SPIE, Bellingham, in press, 2001.

- [12] F. Forbes, C. Fraley, N. Peyrard and A.E. Raftery, "Region-of-interest selection and dynamic breast magnetic resonance data analysis via multivariate and spatial statistical segmentation methods," Technical Report, Department of Statistics, University of Washington, 2001.
- [13] C. Fraley, "Algorithms for model-based Gaussian hierarchical clustering", *SIAM Journal of Scientific Computing*, 20, 270-281, 1999.
- [14] C. Fraley and A.E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis", *The Computer Journal*, 41, 578-588, 1998.



Figure 6: Quantization using ImageMagick. Number of quantization levels used in R, G and B: 9. The original is in color, and this is shown here in greyscale. Dithering is used, yielding a visually better result than no dithering.

- [15] C. Fraley and A.E. Raftery, “Model-Based Clustering, Discriminant Analysis, and Density Estimation,” Technical Report no. 380, Department of Statistics, University of Washington, 2000. Available at www.stat.washington.edu/tech.reports.
- [16] R.E. Kass and A.E. Raftery, “Bayes factors”, *Journal of the American Statistical Association*, 90, 773–795, 1995.
- [17] J. Kominek, Waterloo BragZone, 2000, <http://links.uwaterloo.ca/bragzone.base.html>
Images: <ftp://links.uwaterloo.ca/pub/BragZone/ColorSet>
- [18] G. McLachlan and K. Basford, *Mixture Models*, Marcel Dekker, 1988.
- [19] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, Wiley, 1997.

- [20] MR Multiresolution Analysis Software Environment, Volume 3 – MR/3 Multichannel Data, Section 17.4 Color Images, pp. 287–290, www.multiresolution.com, 2001.
- [21] S. Mukherjee, E.D. Feigelson, G.J. Babu, F. Murtagh, C. Fraley and A. Raftery, “Three types of gamma-ray bursts”, *The Astrophysical Journal*, 508, 314–327, 1998.
- [22] F. Murtagh and A.E. Raftery, “Fitting straight lines to point patterns”, *Pattern Recognition*, 17, 479–483, 1984.
- [23] F. Murtagh, “A survey of algorithms for contiguity-constrained clustering and related problems”, *The Computer Journal*, 28, 82–88, 1985.
- [24] Soo-Chang Pei, Ching-Min Cheng and Lung-Feng Ho, “Limited color display for compressed image and video”, *IEEE Transactions on Circuits and Systems for Video Technology*, 10, 913–922, 2000.
- [25] A.E. Raftery, “Bayesian model selection in social research (with discussion by Andrew Gelman, Donald B. Rubin and Robert M. Hauser)”. In *Sociological Methodology 1995*, Ed. Peter V. Marsden, Oxford: Blackwells, pp. 111–196, 1995.
- [26] K. Roeder and L. Wasserman, “Practical Bayesian density estimation using mixtures of normals”, *Journal of the American Statistical Association*, 92, 894–902, 1997.
- [27] P. Scheunders, “A comparison of clustering algorithms applied to color image quantization”, *Pattern Recognition Letters*, 18, 1379–1384, 1997.
- [28] P. Scheunders, “A genetic c-means clustering algorithm applied to color image quantization”, *Pattern Recognition*, 30, 859–866, 1997.
- [29] G. Schwarz, “Estimating the dimension of a model”, *The Annals of Statistics*, 6, 461–464, 1978.
- [30] D.C. Stanford, Fast Automatic Unsupervised Image Segmentation and Curve Detection in Spatial Point Patterns, Ph.D. Dissertation, Department of Statistics, University of Washington, 1999.
- [31] S. R. Tate, “Band ordering in lossless compression of multispectral images”. *IEEE Transactions on Computers*, 46, 477–483, 1997.

- [32] A. Toet, “Multiscale color image enhancement”, *SPIE International Conference on Image Processing and Its Applications*, SPIE Bellingham, 583–585, 1992.
- [33] D. Tretter, N. Memon and C.A. Bouman, “Multispectral image coding”, in A. Bovik, ed., *Handbook of Image and Video Coding*, Academic, pp. 539–551, 2000.
- [34] F. Truchetet, B. Joanne, F. Péro and O. Laligant, “High-quality still color image compression”, *Optical Engineering*, 39, 409–414, 2000.
- [35] S.J. Wan, S.K.M. Wong and P. Prusinkiewicz, “An algorithm for multidimensional data clustering”, *ACM Transactions on Mathematical Software*, 14, 153–162, 1988. Software: <ftp://ftp.princeton.edu/pub/Graphics/Color/colorquant.shar> (also in Utah Raster Toolkit, see <http://www.cs.utah.edu/gdc/projects/urt>).
- [36] A.B. Watson, G.Y. Yang, J.A. Solomon and J. Villasenor, “Visibility of wavelet quantization noise”, *IEEE Transactions on Image Processing*, 6, 1164–1175, 1997.